Multimodal Deep Generative Models for Trajectory Prediction: A Conditional Variational Autoencoder Approach

Abstract

Human behavior prediction models enable robots to anticipate how humans may react to their actions, and hence are instrumental to devising safe and proactive robot planning algorithms. However, modeling complex interaction dynamics and capturing the possibility of many possible outcomes in such interactive settings is very challenging, which has recently prompted the study of several different approaches. In this work, we provide a self-contained tutorial on a conditional variational autoencoder (CVAE) approach to human behavior prediction which, at its core, can produce a multimodal probability distribution over future human trajectories conditioned on past interactions and candidate robot future actions. Specifically, the goals of this tutorial paper are to review and build a taxonomy of state-of-the-art methods in human behavior prediction, from physics-based to purely data-driven methods, provide a rigorous yet easily accessible description of a data-driven, CVAE-based approach, highlight important design characteristics that make this an attractive model to use in the context of model-based planning for human-robot interactions, and provide important design considerations when using this class of models.